



Heriot-Watt University  
Research Gateway

## Correlation analysis of air pollutant index levels and dengue cases across five different zones in Selangor, Malaysia

### Citation for published version:

Thiruchelvam, L, Dass, SC, Zaki, R, Yahya, A & Asirvadam, VS 2018, 'Correlation analysis of air pollutant index levels and dengue cases across five different zones in Selangor, Malaysia', *Geospatial Health*, vol. 13, no. 1, pp. 102-109. <https://doi.org/10.4081/gh.2018.613>

### Digital Object Identifier (DOI):

[10.4081/gh.2018.613](https://doi.org/10.4081/gh.2018.613)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

Geospatial Health

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Correlation analysis of air pollutant index levels and dengue cases across five different zones in Selangor, Malaysia

Loshini Thiruchelvam,<sup>1</sup> Sarat C. Dass,<sup>1</sup> Rafdzah Zaki,<sup>2,3</sup> Abqariyah Yahya,<sup>2,3</sup> Vijanth S. Asirvadam<sup>4</sup>

<sup>1</sup>*Fundamental and Applied Sciences Department, Faculty of Science and Information Technology, Universiti Teknologi Petronas, Bandar Seri Iskandar, Tronoh, Perak;* <sup>2</sup>*Julius Centre University of Malaya, Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur;* <sup>3</sup>*Public Health Department, University of Malaya Medical Centre, Kuala Lumpur;* <sup>4</sup>*Department of Electrical and Electronic Engineering, Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi Petronas, Bandar Seri Iskandar, Tronoh, Perak, Malaysia*

## Abstract

This study investigated the potential relationship between dengue cases and air quality – as measured by the Air Pollution Index (API) for five zones in the state of Selangor, Malaysia. Dengue case patterns can be learned using prediction models based on feedback (lagged terms). However, the question whether

air quality affects dengue cases is still not thoroughly investigated based on such feedback models. This work developed dengue prediction models using the autoregressive integrated moving average (ARIMA) and ARIMA with an exogenous variable (ARIMAX) time series methodologies with API as the exogenous variable. The Box Jenkins approach based on maximum likelihood was used for analysis as it gives effective model estimates and prediction. Three stages of model comparison were carried out for each zone: first with ARIMA models without API, then ARIMAX models with API data from the API station for that zone and finally, ARIMAX models with API data from the zone and spatially neighbouring zones. Bayesian Information Criterion (BIC) gives goodness-of-fit versus parsimony comparisons between all elicited models. Our study found that ARIMA models, with the lowest BIC value, outperformed the rest in all five zones. The BIC values for the zone of Kuala Selangor were –800.66, –796.22, and –790.5229, respectively, for ARIMA only, ARIMAX with single API component and ARIMAX with API components from its zone and spatially neighbouring zones. Therefore, we concluded that API levels, either temporally for each zone or spatio-temporally based on neighbouring zones, do not have a significant effect on dengue cases.

Correspondence: Sarat C. Dass, Fundamental and Applied Sciences Department, Faculty of Science and Information Technology, Universiti Teknologi Petronas, 32610 Bandar Seri Iskandar, Tronoh, Perak, Malaysia.

Tel.: +605.368.7038 - Fax: +605.365.5905.

E-mail: saratcdass@gmail.com

Key words: Dengue; Air pollutant index; Bayesian information criterion value.

Acknowledgements: authors would like to thank the Ministry of Health and Department of Environment, Malaysia for providing the necessary data to conduct this work.

Contributions: the authors contributed equally.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: this research was partly supported by the University of Malaya Research Grant Programme [RP034A-15HTM] and by the Fundamental Research Grant Scheme (FRGS-0153AB-L19), from the Ministry of Education, Malaysia.

Received for publication: 7 August 2017.

Revision received: 25 August 2017.

Accepted for publication: 15 November 2017.

©Copyright L. Thiruchelvam et al., 2018

Licensee PAGEPress, Italy

Geospatial Health 2018; 13:613

doi:10.4081/gh.2018.613

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Introduction

Dengue fever, which is transmitted by *Aedes* mosquitoes, have been increasing in Malaysia lately, specifically in its most urbanized state, Selangor (Hassan *et al.*, 2012). Uncontrolled population growth, climate change and unplanned urbanization are some of the main factors linked to an increase in dengue cases (Hassan *et al.*, 2012). Control against the vector population is still the best way to curb this disease as vaccination is still under study with mixed outcomes. A dengue vaccine named *Dengvaxia* has been developed by Sanofi Pasteur Ltd (Sanofi Pasteur Press Release, 2016), but this vaccine has caused many young children who underwent vaccination to become more sick after a three-year time period (Warner, 2016). As a result of this, dengue prediction models are still important and complementary tools for controlling the vector population at present.

Most of the dengue prediction models use meteorological parameters such as mean temperature, relative humidity and rainfall as predictors (Brunkard *et al.*, 2008; Olson *et al.*, 2009; Cheong *et al.*, 2012; Cheong *et al.*, 2013; Dom *et al.*, 2013; Naish

*et al.*, 2014). Inclusion of air quality as a potential predictor is still a topic that is not well researched in the literature; there are only a few reported studies, especially from Singapore, which have addressed this issue (Smith *et al.*, 2013). The name of measurement that refers to air quality also varies among countries. For example, it is known as the Air Pollution Index (API) in Malaysia, as Air Quality Index (AQI) in Thailand and as Pollutant Standard Index (PSI) in Singapore (ASEAN Specialised Meteorological Centre, 2017). A previous study (Massad *et al.*, 2010) based in Singapore hypothesized that a severe haze condition (with high PSI levels) causes a drop in dengue since severe haze will produce smoke that would kill the adult mosquitoes. The authors' discussion was based on observations of drops in dengue cases during the Southeast Asia extreme haze event in 2006. On the other hand, Smith *et al.* (2010) carried out a haze-dengue analysis, also in Singapore, and concluded oppositely: High PSI value did not necessarily reduce dengue cases. They observed that during the bad haze conditions in years of 1997 and 1998, dengue cases were also at its highest levels. Therefore, they concluded that severe haze conditions did not necessarily bring about a drop in dengue cases. These two opposite findings can be further explained by another study by Oliver *et al.* (2014) which have discussed that only when the API level is enough high, its negative effect on the mosquito population (these vectors could be destroyed) can be felt and dengue occurrences could be reduced (Oliver *et al.*, 2014). Therefore, the relationship between air quality levels and dengue cases remains uncertain.

In view of this, our study aimed to explore API as a potential predictor of dengue cases for zones involved in our study in Malaysia. A challenge in this study is that previous dengue cases provided a positive feedback for current dengue cases due to the nature of transmission of this disease. Thus, the effect of API had to be assessed over and above this feedback effect. To address whether API levels correlate with dengue cases, a number of autoregressive integrated moving average (ARIMA) and autoregressive integrated moving average with exogenous variable (ARIMAX) models were built using the Box-Jenkins (BJ) variable transformation approach. Subsequently, the ARIMA and ARIMAX models were optimized with respect to the extent of their feedback. The optimized ARIMA and ARIMAX models were then compared using the Bayesian Information Criteria (BIC) as the criteria for model selection. Our approach of deciding the presence or absence of correlation of API with dengue cases was based on three stages of model-based analysis: model identification, validation and selection. Our models were able to find long-term correlations between API levels and dengue cases, if any, after accounting for other explanatory attributes proxied through lagged dengue cases.

## Materials and Methods

### Study area

This study involved the following five zones in Selangor: i) Petaling Jaya; ii) Pelabuhan Klang; iii) Shah Alam; iv) Kuala Selangor; v) Banting. The state of Selangor was selected as it recorded the highest number of dengue cases in Malaysia (Hassan *et al.*, 2012). Selangor is also well known for facing deterioration of air quality every year, where it recorded a number of 59, 48, and 37 unhealthy days (API levels were in the range of 101-200) during the years of 2009-2012. This poor air quality is believed to

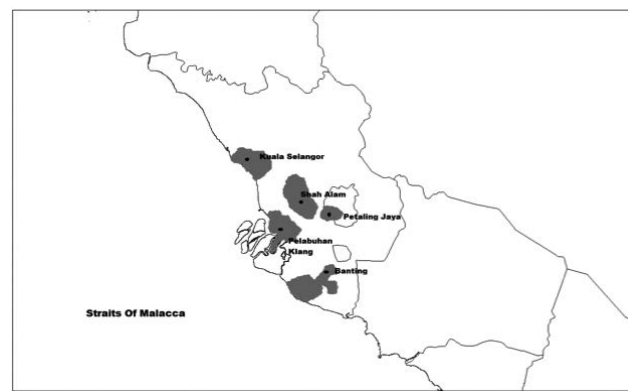
occur due to a number of factors such as commercial and industrial developments, emissions from motor vehicles and the trans boundary haze (Rahmah *et al.*, 2015).

Data on weekly dengue cases were obtained from the Ministry of Health Malaysia, and it represents the total number of dengue cases that occurred in the respective zones. On the other hand, data on API measurements were obtained from the Department of Environment, Malaysia. The API measurements for Malaysia comprise measurements of five main air pollutants, *i.e.* carbon monoxide (CO), ozone (O<sub>3</sub>), nitrogen dioxide (NO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>) and particulate matter with a diameter of less than 10 micron (PM<sub>10</sub>) (Department of Environment Malaysia, 2017). All dengue cases and API data for the five zones of study were reported on a weekly basis, starting from the year 2009 till 2015, except for Banting where the dengue cases and API data were obtained from year 2010 till 2015. The surrounding environment for air quality monitoring stations varied: the stations in Petaling Jaya, Shah Alam and Pelabuhan Klang are located in the middle with main roads which connect industrial and high density housing areas. The Kuala Selangor station is located in a comparatively less populated area, whereas the Banting station is near the roadside of a highway, thus located in an area that is less populated and having trees and oil palm plantations (Rahmah *et al.*, 2015). These different environments contribute to the variation in levels of deterioration of air quality measurements (Azmi *et al.*, 2010). Figure 1 shows the spatial locations of the API monitoring stations as well as the boundaries of the zones considered. The areas corresponding to these zones are as follows: Petaling Jaya (PJ) = 97.2 km<sup>2</sup>, Pelabuhan Klang (PK) = 261 km<sup>2</sup>, Shah Alam (SA) = 293 km<sup>2</sup>, Kuala Selangor = 217 km<sup>2</sup> and Banting = 261 km<sup>2</sup>. Four of the areas are approximately similar but PJ is half the size of the rest.

The locations of API stations for the 5 zones, Petaling Jaya, Pelabuhan Klang, Shah Alam, Kuala Selangor and Banting, are marked with black dots and their zonal boundaries indicated by the shaded grey patches.

### Presence of feedback

It is proposed that dengue cases can be explained through various time-lagged models with lags of up to ten weeks. The extent of feedback can be understood based on the study of the lifecycle of the *A. aegypti* mosquito. The mosquito breeding period from mosquito eggs to adulthood (Life cycle of *Aedes Aegypti*, 2017) lasts between 1.5 to 3 weeks, which is followed by a biting period where the adult mosquitoes, if successful in biting an infected



**Figure 1. Map outlining the boundaries of the state of Selangor in West Malaysia.**

human, will acquire the dengue virus. The infected mosquito then undergoes an extrinsic incubation period (EIP) which takes from 8 to 12 days before becoming infectious (Chen *et al.*, 2012). This is followed by a mosquito transmission period where the infected mosquitoes will bite susceptible (uninfected) humans (Dengue Transmission, 2014). The next phase is the human transmission phase where the dengue virus is transferred from an infected person to an uninfected person (Cheong *et al.* 2013). The virus will undergo an Intrinsic Incubation Period (IIP) inside the newly infected human which lasts for 4-10 days after which the infected person will exhibit symptoms of dengue (World Health Organization, 2009). The number of infected humans at previous time points, therefore, serves as a determinant of current levels of dengue cases based on the lifecycle of *Aedes* mosquitoes as described above. As the number of infected humans at any time point is difficult to determine, a reliable substitute is the actual number of dengue cases observed at these previous time points. Variability in the biting period, the mosquito transmission period, human transmission period, the EIP and the IIP cause us to consider all previous dengue cases up to lag-10 weeks as potential predictors for the current number of dengue cases, which is the maximum extent of feedback that we consider in our ARIMA and ARIMAX models. Figure 2 shows the timeframe of the dengue infection process. For dengue cases, ARIMA models are considered as models that can incorporate feedback information via lagged terms. In this study, API is the exogenous variable considered in our ARIMAX models. Both models are used to represent stationary time series data (Alain *et al.*, 2007; Cressie and Wikle, 2011). However, when a time series is non-stationary, as is the case for the dengue case trajectories, ARIMA and ARIMAX models cannot be applied directly. The standard practice is to have the original time series differenced up to the first or second order to yield stationarity after which the models can be fit to the observed data (Hyndman and Athanasopoulos, 2013).

## Statistics

In our case, first order differencing was found to be sufficient to yield stationarity; higher order differencing was found to be unnecessary. The general mathematical equation of an ARIMA model for a differenced time series  $Y^*$  is:

$$Y_t^* = \phi_1 Y_{t-1}^* + \phi_2 Y_{t-2}^* + \dots + \phi_p Y_{t-p}^* + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad \text{Eq. 1}$$

where  $\phi = (\phi_1, \phi_2, \dots, \phi_p)$  and  $\theta = (\theta_1, \theta_2, \dots, \theta_q)$  are unknown coefficients with each  $\phi_i \in \mathbb{R}$  and  $\theta_j \in \mathbb{R}$ , and  $\varepsilon$ 's random errors independent and identically distributed with zero mean and constant

but unknown variance,  $\sigma^2$ . In Eq. 1,  $q$  is the extent of lag (*i.e.* the extent of dengue case feedback) and the extent of lag of the errors (*i.e.* the extent of correlation in  $Y_t^*$ ). The ARIMA model in Eq. 1 with lag orders  $p$  and  $q$  is denoted ARIMA( $p, d, q$ ) where  $d$  is the order of differencing to yield stationarity. The values of  $p$  and  $q$  were obtained based on inspecting the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of  $Y_t^*$ . The method of choosing  $p$  and  $q$  based on inspection is a well-known technique in the literature (see Nau, 2014) and hence, we omit further details here. To avoid complicated models, the  $\phi$ - and  $\theta$ -coefficients in Eq. 1 corresponding to insignificant lags in the ACF and PACF plots were set to zero. The final model was denoted by ARIMA( $p, d, q$ )\* where \* indicates the reduced model with omission of insignificant coefficients.

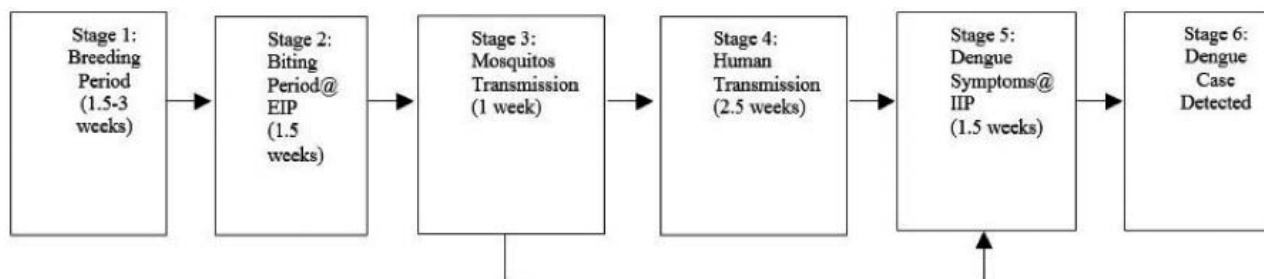
The remaining significant but unknown coefficients in the ARIMA( $p, d, q$ )\* model were estimated using the maximum likelihood (ML) approach (Wang and Jain, 2003; Levine and Justice, 2013). The ML approach was used as it reports good model fit and prediction accuracy. Subsequently, residual diagnostics based on the fitted model was performed. The fitted model is deemed appropriate if the ACF and PACF plots of the residuals do not show any significant autocorrelations (*i.e.* significant peaks at lagged time points). If not, the above process must be repeated for different choices of  $p$  and  $q$  until desired results are obtained. We provide an illustration of the techniques for one of the zones under study in the Results section. The details on the same methodology are omitted in the analysis corresponding to the other four zones.

Exogenous variables can be incorporated into the ARIMA( $p, d, q$ )\* model forming an ARIMAX( $p, d, q, m$ )\* model with  $m$  lags. The corresponding mathematical equation of an ARIMAX( $p, d, q, m$ )\* model is:

$$Y_t^* = \sum_{i=0}^m \beta_i X_{t-i} + \phi_1 Y_{t-1}^* + \phi_2 Y_{t-2}^* + \dots + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad \text{Eq. 2}$$

where  $\sum_{i=0}^m \beta_i X_{t-i}$  is the linear combination of lagged API observations

up to order  $m$ . The  $\phi$ - and  $\theta$ -coefficients which were insignificant in the original ARIMA( $p, d, q$ )\* model were set to zero in the ARIMAX( $p, d, q, m$ )\* model in Eq. 2 as well. The remaining significant  $\phi$ - and  $\theta$ -coefficients as well as the  $m+1$   $\beta$ -coefficients in the ARIMAX( $p, d, q, m$ )\* model in Eq. 2 were re-estimated using ML. We noted that the API time series also needs to be stationary to avoid spurious regression (Hyndman and Athanasopoulos, 2013) which was achieved by performing first order differencing as mentioned earlier.



**Figure 2.** Life cycle of the *Aedes Aegypti* mosquito and dengue transmission to humans: a motivation for lagged time series models for dengue cases.



We selected an upper bound for  $m$ , namely  $m_0$ , and fit  $m_0$  ARIMAX( $p, d, q, m$ )\* models for  $m=0, 1, 2, \dots, m_0$ . Among the ARIMA( $p, d, q$ )\* and the  $m_0$  ARIMAX( $p, d, q, m$ )\* models, the best one was selected using (BIC):

$$BIC = -2(\log L) + k * \log(n) \quad \text{Eq. 3}$$

where  $L$  is the model likelihood,  $k$  is the number of model parameters and  $n$  the number of observations in the time series data. Model selection based on BIC accounts for both model fit and model parsimony (Bhatnagar *et al.*, 2012; Sitepu *et al.*, 2013; Sriyasatien *et al.*, 2016). In Eq. 3,  $k = k_0 + 1$  where  $k_0$  is the number of significant  $\phi$ - and  $\theta$ -coefficients in the ARIMA( $p, d, q$ )\* model; the extra +1 corresponds to the variance parameter  $\sigma^2$ . For the ARIMAX( $p, d, q, m$ )\* model, the value of  $k = k_0 + m + 2$  since there are  $m + 1$   $\beta$ -coefficients and one variance parameter  $\sigma^2$ . The best model is the one with the lowest BIC value in Eq. 3. We also considered spatio-temporal analysis where temporal API measurements from spatially neighbouring zones were incorporated into Eq. 2 in addition to the API measurements from that zone. In this case, the value of  $k = k_0 + \sum_{j=1}^M m_j + M + 1$  in Eq. 3 where  $M$  is the total number of zones considered with  $m_j$  denoting the lag order for zone  $j = 1, 2, \dots, M$ .

## Results

### Description of data

A total of 354 weekly data were used (covering the period between January 2009 up to middle of October 2015) involving observations of dengue cases and API levels; for Banting, the number of weeks covered was 290 from March 2010 to October 2015. Table 1 gives the range and average values of the data for each zone. The average API readings ranged between 51 and 100 which is classified as moderate, that is, *moderate pollution that does not pose any bad effect on health* (Department of Environment

Malaysia, 2017). This moderate API range should be acceptable for most of the public except for a small group of people who are sensitive to ozone or particulate pollution (Department of Environment Malaysia, 2017).

### Prediction models

Prior to model building and selection, we normalized both the response and explanatory variables (*i.e.* dengue cases and API levels) into the same range (0-1): Letting  $H$  represent a generic notation for a variable, normalization entails

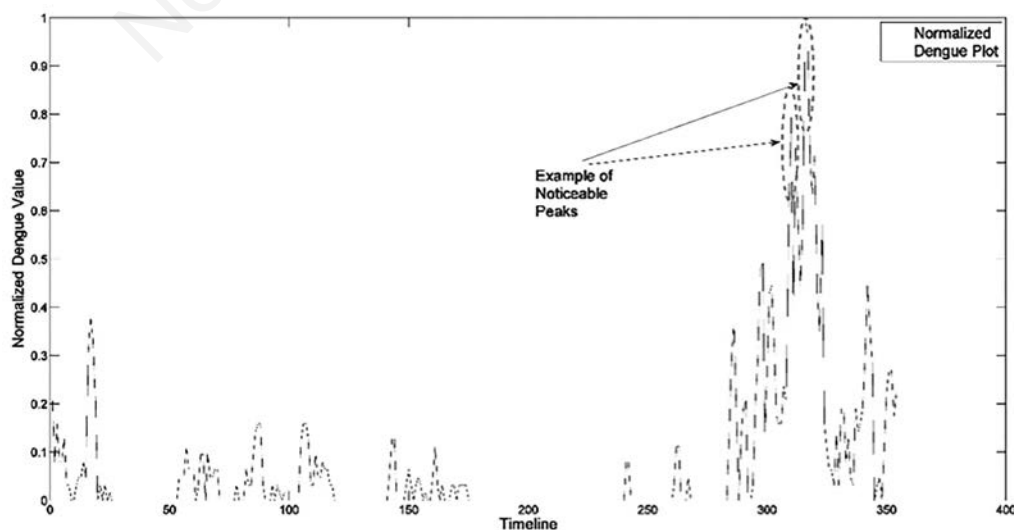
$$H_N = \frac{H - H_{\min}}{H_{\max} - H_{\min}} \quad \text{Eq. 4}$$

where  $H_N$  is the normalized variable, and  $H_{\min}$  and  $H_{\max}$ , the minimum and maximum values of  $H$ , respectively, in the data set.

Three models were compared in this study, *i.e.* the ARIMA models, ARIMAX models with single-API station and the ARIMAX with paired-API stations from spatially neighboring zones. To build the ARIMA model, the BJ approach was used: first, differencing was carried out to ensure that the data were stationary; next, the significant lag orders of  $p$  and  $q$  were selected by inspecting the ACF and PACF plots; finally, the ARIMA model was estimated to obtain the parameter values using the ML approach. An example of ARIMA model building is illustrated for the Kuala Selangor zone in Figures 3-5. Figure 3 shows the weekly dengue

**Table 1. Summary measures of air pollutant index (API) levels and dengue cases.**

Area	API levels		Dengue cases	
	Average	Range	Average	Range
Petaling Jaya	58.61	[34.46, 201.29]	267.92	[0, 1585.00]
Pelabuhan Klang	70.24	[37.00, 450.13]	57.48	[0, 522.00]
Shah Alam	61.53	[34.33, 248.29]	213.27	[0, 1782.00]
Kuala Selangor	55.31	[23.58, 214.75]	4.71	[0, 63.00]
Banting	63.69	[40.00, 264.54]	8.63	[0, 89.00]



**Figure 3. Plot of dengue cases for Kuala Selangor region.**

case trajectory for Kuala Selangor during the years 2009-2015 which is clearly seen to be non-stationary. First order differencing ( $d=1$ ) was carried out and significant lag orders identified using the ACF and PACF plots. The ACF plot in Figure 4 shows a significant peak at  $q=2$ , hence, the ARIMA  $(0, 1, 2)^*$  model was selected. We found that significant peaks in both the ACF and PACF plots can be eliminated by choosing exactly one of  $p$  or  $q$  appropriately, that is, by setting the other to zero, which also reduces model complexity. Figure 5 shows the ACF and PACF plots of the residuals from the fitted model. No significant peaks existed outside the 95% confidence lines indicating that the residuals are indeed white noise. Hence, the ARIMA model  $(0, 1, 2)^*$  was selected as the model for dengue cases for Kuala Selangor.

To investigate if API had any effect on dengue cases, six ARI-

MAX $(0,1,2,m)^*$  models were formed for different lag orders  $m=0, 1, 2, \dots, m_0=5$ . Table 2 gives the BIC values for ARIMA and best ARIMAX models (out of the six, based on BIC) for Kuala Selangor. A similar methodology was also applied to the rest of the study zones (Table 2).

Next, a spatio-temporal analysis was carried out to find if inclusion of API parameter from spatially neighbouring stations (the closest one) could enhance the model fit. Table 3 shows the results obtained for the ARIMAX model for Kuala Selangor when API data from its closest neighbour, Shah Alam, was additionally incorporated into the ARIMAX model (Figure 1). Results from Table 3 show that best ARIMAX model is from the pairing of API parameters at lag orders of zero and zero. A comparison of the BIC values in Tables 3 and 4 shows that it reaches its lowest value

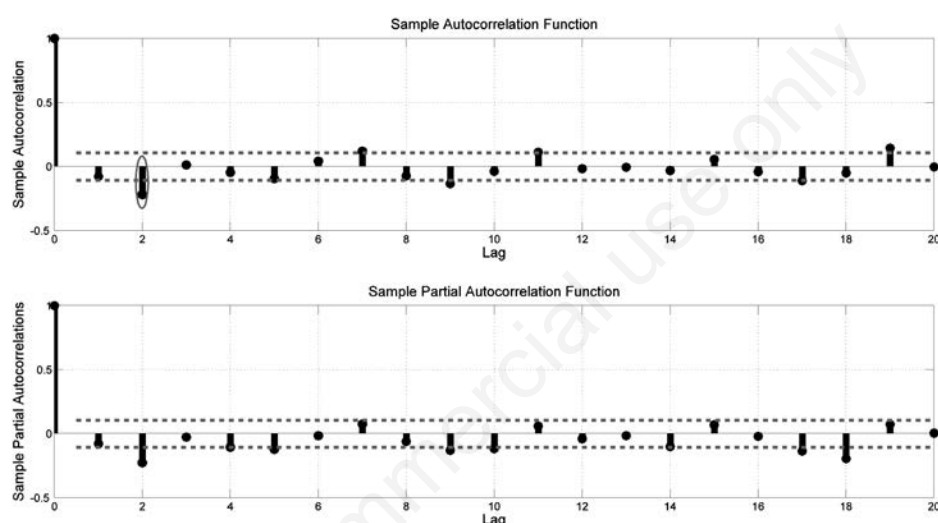


Figure 4. Plot of autocorrelation function and partial autocorrelation function after differencing.

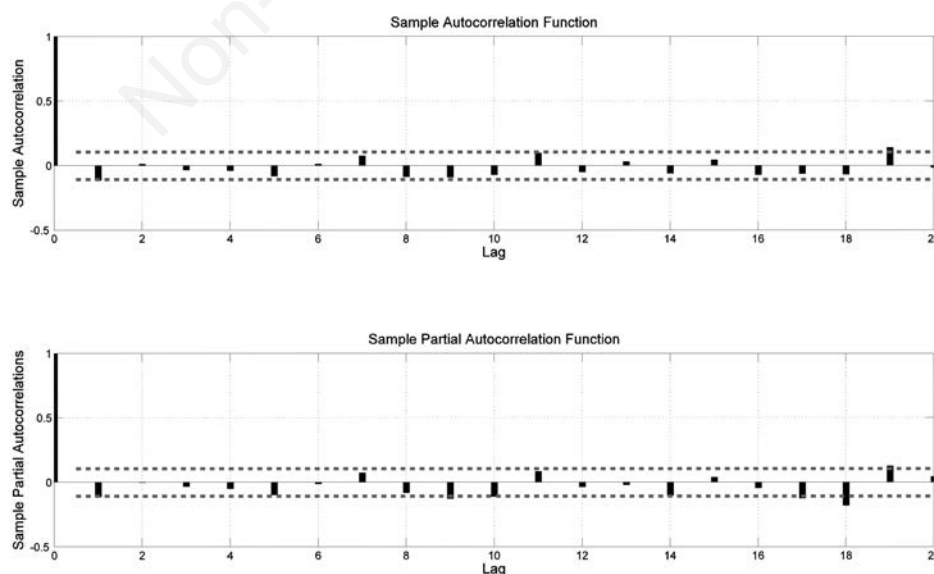


Figure 5. Plot of autocorrelation function and partial autocorrelation function for residuals of the autoregressive integrated moving average  $(0,1,2)$  model.

**Table 2. Best model for each study zone selected using Bayesian information criterion (BIC).**

Study zone	Best ARIMA model	BIC value	Best ARIMAX model	BIC value
Petaling Jaya	ARIMA (7,1,9)	-964.5606	ARIMAX (7,1,9,0)	-958.7241
Shah Alam	ARIMA (2,1,1)	-1.0075e+03	ARIMAX (2,1,1,0)	-1.0030e+03
Pelabuhan Klang	ARIMA (3,1,1)	-858.3920	ARIMAX (3,1,1,0)	-853.7578
Kuala Selangor	ARIMA (0,1,2)	-800.6577	ARIMAX (0,1,2,0)	-796.2189
Banting	ARIMA (0,1,3)	-519.0362	ARIMAX (0,1,3,0)	-516.1652

ARIMA, Autoregressive integrated moving average, ARIMAX, ARIMA with an exogeneous variable.

for the ARIMA model without any API component. We concluded that the dengue trajectory for Kuala Selangor was very localized and intrinsic to that zone only. Once the localized, best model based on feedback had been identified, it was realized that API observations for that zone as well as API data from spatially neighbouring zones do not contribute to better models. Indeed, similar results were found for all other zones: Comparing the best ARIMA and ARIMAX models from Table 2 with the best ARIMAX models from Table 4 (with API data from spatially neighbouring zones), we found that the ARIMA models had best BIC values in all the five zones of study.

## Discussion

In Malaysia, poor air quality mostly occurs during the haze episodes, known as the trans-boundary haze (see BBC News Online, 2017) due to smoke from open burning and wildfires in Indonesia. Haze and other local particulate matters were believed to kill the mosquito vectors, causing lower dengue transmission and decreased number of dengue incidences as a whole (Massad *et al.*, 2010). Indeed, when their modelled prediction of a dengue outbreak did not materialize, Massad *et al.* (2010) suggested that haze could have prevented it. This suggestion was based on Singapore data on dengue and the pollutant standards index (PSI) for the year of 2006 and 2007. Smoke is also known to act as an insect repellent and therefore used as such (Tabuti, 2007).

Studies which investigate the relationship between haze (or air quality index) and the number of dengue cases in Malaysia are at best scarce. Rafdzah *et al.* (2015) conducted an extensive review of existing literature for such studies and came up with a list that does not contain any studies from Malaysia. This motivated us to conduct this study in the state of Selangor in Malaysia. Based on our analysis, we found no relationship between API level and dengue cases in all five study zones in Selangor. Prediction models were better elicited without API as a predictor. Based on data from the region of Kuala Selangor, we found that the optimum ARIMA model was better than the optimal ARIMAX model based on BIC values. The reason was that ARIMAX( $p, d, q, m$ )\* models are more complex compared to ARIMA( $p, d, q$ )\* and will be penalized by the BIC criteria (*i.e.* result in higher BIC values) if the inclusion of API as the exogeneous variable is redundant, that is, if API has no effect on dengue cases. A spatio-temporal analysis also yielded similar conclusions. Incorporation of API measurements from Shah Alam, the station closest to Kuala Selangor, in addition to API measurements from Kuala Selangor station did not lower the BIC values, which means that their incorporation into the model was redundant. Similar results were observed for the four other study areas.

**Table 3. Best Bayesian information criterion values for Kuala Selangor autoregressive integrated moving average with an exogeneous variable models for dengue cases with air pollutant index measurements from this zone and its closest spatial neighbouring zone, ShahAlam.**

Pairing station Shah Alam	Kuala Selangor BIC values for models at		
	Lag 0	Lag 1	Lag 2
Lag 0	-790.5229	-786.1049	-778.9426
Lag 1	-785.9545	-780.2422	-773.0930
Lag 2	-779.3796	-773.7084	-767.9281

BIC, Bayesian information criterion.

**Table 4. Best Bayesian information criterion (BIC) values for each zone with respect to their nearest spatial neighbours.**

Study Zone	Closest spatial neighbour(s)	Best BIC value(s)
Petaling Jaya	Shah Alam	-952.92
Shah Alam	Pelabuhan Klang, Petaling Jaya	-998.25, -997.33
Pelabuhan Klang	Shah Alam, Kuala Selangor	-848.68, -848.01
Kuala Selangor	Shah Alam	-790.52
Banting	Pelabuhan Klang, Petaling Jaya, Shah Alam	-511.61, -511.04, -511.82

Dengue case observations were similar in the closely situated neighbouring zones Petaling Jaya and Shah Alam as can be seen in Table 1. The average and range of dengue cases were 267.92 [0, 1585.00] for Petaling Jaya and 213.27 [0, 1782.00] for Shah Alam, respectively. These similar average and range values differed significantly from the other zones in Table 1. In terms of API levels, there were similarities too, *i.e.* in the average and range of API levels. As seen in Table 1, the average and range of API levels were 58.61 [34.46, 201.29] for Petaling Jaya and 61.53 [34.33, 248.29] for Shah Alam, respectively. These average and range values again differed significantly compared to the other zones. Due to this similarity between Petaling Jaya and Shah Alam, the best models (which were the ARIMA models although with differing ARIMA components) had BIC values closer to each other compared to the other zones (Table 2), meaning that these models provided similar model fit and parsimonious assessment to be used for prediction purposes.

Our study findings on dengue cases and API levels are consistent with findings from a previous study (Smith *et al.*, 2010) based on 2001-2008 data from Singapore and air quality measured according to the PSI. They concluded that there was no significant association between PSI and dengue. However, their data were



limited to a small range of dengue cases (minimum of 45 and maximum of 265) and PSI (minimum of 34.2 and maximum of 41.9) (Rafdzah *et al.*, 2015). We considered a higher range of dengue cases as well as API values in this paper and still did not find any correlation among them.

Interestingly, some positive impact of haze on dengue has been suggested (Ooi *et al.*, 2010; Smith *et al.*, 2013) based on the 1997–1998 dengue outbreak in Singapore, which faced severe haze at that time. This impact was explained in Smith *et al.* (2013) where they suggested the high haze readings have caused both human and mosquitoes to stay indoors bringing higher possibilities for mosquito bites and dengue transmission. However, Massad *et al.* (2010) suggested that the observed outbreak during this period could be explained by the high annual average temperature which occurred together with the haze conditions. Annual average temperatures were at the highest during the latest 15 years at the time, and this could have caused an increase in the number of infectious *Aedes* mosquitoes since the EIP period would have been shortened. This, in turn, would have caused more possibilities for infectious bites and dengue transmission. Thus, in our model, it was imperative to include such explanations in the model development so that the effect of API levels could be assessed after such inclusions; note that our model used time-lagged dengue cases as a proxy for high temperature. Thus, residual correlation of API levels and dengue cases, if any, would only be reflected in our models if truly present and not due to other confounding factors such as high temperature. Indeed, we found that after the inclusion of time-lagged dengue cases, API levels did not provide any additional explanation to the increase or decrease in the number of dengue cases.

## Conclusions

This is one of the first studies to investigate the relationship between API levels and the number of dengue cases in Malaysia using feedback modelling techniques such as ARIMA and ARIMAX. Our study found that API level did not correlate with dengue cases in all five study zones in the state of Selangor, Malaysia, which were selected due to their high levels of dengue cases. The inclusion of API for all study zones did not enhance our prediction models as assessed by the BIC values. The BIC values did not improve, either by adding API data from spatially neighbouring zones. We conclude that the dengue cases trajectory for each zone was very localized and intrinsic to that zone only. Once the localized best model was identified; other variables (the API observations for the focal zone as well as API data from spatially neighbouring zones) did not contribute to yield better prediction models. The finding that the ARIMA models with the lowest BIC values outperformed the rest of the models in all five zones should be of future interest.

## References

- Alain FZ, Elena NL, Graham MS, 2007. Time series analysis. In: Zurr AF, Ieno EN, Smith GM, eds. *Analyzing ecological data*. New York, NY: Springer; pp 264–288.
- Anno S, Imaoka K, Tadono T, Igarashi T, Sivaganesh S, Kannathan S, Kumaran V, Surendran SN, 2015. Space-time clustering characteristics of dengue based on ecological, socio-economic and demographic factors in northern Sri Lanka. *Geospat Health* 10:215–8.
- Azmi SZ, Latif MT, Ismail AS, Juneng L, Jemain AA, 2010. Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. *Air Qual Atmosphere Health* 3:53–12.
- ASEAN Specialised Meteorological Centre (ASMC). Air Qual 2017; Available from: <http://asmc.asean.org/asmc-haze-air-quality/>. Accessed: October 2017.
- BBC News Online. What causes South East Asia's haze? Available from: <http://www.bbc.com/news/world-asia-34265922>. Accessed: May 2017.
- Bhatnagar S, Lai V, Gupta S, 2012. Forecasting incidence of dengue in Rajasthan, using time series analyses. *Indian J Pub Health* 56:281–5.
- Brunkard JM, Cifuentes E, Rothenberg SJ, 2008. Assessing the roles of temperature, precipitation, and ENSO in dengue re-emergence on the Texas-Mexico border region. *Salud Pública México* 50:227–9.
- Chen MJ, Lin CY, Wu YT, 2012. Effects of extreme precipitation to the distribution of infectious diseases in Taiwan. *PLoS One* 7:1994–16.
- Cheong YL, Burkart L, Leitao PJ, Lakes T, 2013. Assessing weather effects on dengue disease in Malaysia. *Int J Environ Res Pub Health* 10:6319–6.
- Cressie N, Wikle CK, 2011. *Statistics for spatio-temporal data*. New Jersey, USA: Wiley Series in Probability and Statistics.
- Department of Environment Malaysia. Air Pollution Index of Malaysia. 2017; Available from: <http://apims.doe.gov.my/v2/>. Accessed: January 2017.
- Dengue Transmission, 2014. Available from: <http://www.nature.com/scitable/topicpage/dengue-transmission-22399758>. Accessed: January 2017.
- Dom NC, Hassan AA, Latif ZA, Ismail R, 2013. Generating temporal model using climate variables for the prediction of dengue cases in Subang Jaya, Malaysia. *Asian Pacific J Trop Dis* 3:352–61.
- Evelyn Warner, 2016. Too good to be true: Sanofi Pasteur's dengue vaccine raises serious concerns. Available from: <http://labiotech.eu/dengvaxia-sanofi-pasteur-dengue-vaccine/>. Accessed: May 2017.
- Hassan H, Shohaimi S, Hashim NR, 2012. Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia. *Geospat Health* 7:21–5.
- Hyndman RJ, Athanasopoulos G, 2013. *Forecasting: principles and practice*. OTexts: Melbourne, Australia. Available from: <http://otexts.org/fpp/>. Accessed: March 2017.
- Levine H, Justice JB, 2013. *Handbook of local government fiscal health*. Burlington, MA, USA: Jones & Bartlett Learning.
- Life cycle of *Aedes Aegypti*. Available from: <http://www.dengue-virusnet.com/>. Accessed: January 2017.
- Massad E, Coutinho F, Ma S, Burattini M, 2010. A hypothesis for the 2007 dengue outbreak in Singapore. *Epidemiol Infect* 138:951–7.
- Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S, 2014. Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC Infect Dis* 14:167.
- Nau R, 2014. Notes on nonseasonal ARIMA models. Available from: [http://people.duke.edu/~rnau/Notes\\_on\\_nonseasonal\\_](http://people.duke.edu/~rnau/Notes_on_nonseasonal_)





- ARIMA\_models—Robert\_Nau.pdf. Accessed: January 2017.
- Ooi EE, Wilder-Smith A, Ng LC, Gubler DJ, 2010. The 2007 dengue outbreak in Singapore. *Epidemiol Infect* 138:958-61.
- Oliver L, Burattini MN, Coutinho FA, Coelho GE, Struchiner CJ, Massad E, 2014. A negative correlation between dengue and bushfires in Brazil. *J Environ Health* 76:66-7.
- Olson J, Berggren K, Olofsson M, Viklander M, 2009. Applying climate model precipitation scenarios for urban hydrological assessment: A case study in Kalmar City, Sweden. *Atmosph Res* 92:364-75.
- Rafdzah Z, Nasrin A, Kavita R, Abqariah Y, 2015. The relationship between air quality and dengue cases: a systematic review. Available from: <http://apacph2015.fkm.ui.ac.id/ppt/21%20October%202015/9.%20FP%20Environmental%20Health%20C-Amartapura%20C/8.%20Rafdzah%20Zaki.pdf>. Accessed: January 2017.
- Rahmah S, Norkhadajah S, Ramli MF, Latif MT, 2015. The assessment of ambient air pollution trend in Klang Valley, Malaysia. *World Environment* 5:1-11.
- Sanofi Pasteur Press Release, 2016. Available from: <http://www.sanofipasteur.com/en/articles/New-England-Journal-of-Medicine-Publishes-New-Analyses-Confirming-that-Sanofi-Pasteur-sVaccine-Candidate-Safely-Protects-Pre-Adolescents-to-Adults-Against-Dengue.aspx>. Accessed: August 2016.
- Sarfraz MS, Tripathi NK, Faruque FS, Bajwa UI, Kitamoto A, Souris M, 2014. Mapping urban and peri-urban breeding habitats of *Aedes* mosquitoes using a fuzzy analytical hierarchical process based on climatic and physical parameters. *Geospat Health* 8:685-97.
- Sitepu M, Kaewkungwal J, Luplerdlop N, 2013. Temporal patterns and a disease forecasting model of dengue hemorrhagic fever in Jakarta based on 10 years of surveillance data. *Southeast Asian J Trop Med Pub Health* 44:206-17.
- Sriyasatien P, Phumee A, Ongkruk P, 2016. Analysis of significant factors for dengue fever incidence prediction. *BMC Bioinformatics* 17:166-9.
- Tabuti JRS, 2007. Herbal medicines used in the treatment of malaria in Budiope country, Uganda. *J Ethnopharmacol* 116:33-42.
- U.S. Environmental Protection Agency. A guide to air quality and your health, 2014. Available from: [https://airnow.gov/index.cfm?action=aqi\\_brochure.index](https://airnow.gov/index.cfm?action=aqi_brochure.index). Accessed: July 2017.
- Wang GCS, Jain CL. Regression analysis: modelling & forecasting. Flushing, New York: Graceway Publishing Company Inc.; 2003.
- Wilder-Smith A, Earnest A, Tan S, Ooi E, Gubler D, 2010. Lack of association of dengue activity with haze. *Epidemiol Infect* 138:962-6.
- Wilder-Smith A, Earnest A, Ma S, Ng LC, 2013. Haze and dengue: the unanswered questions. *Ann Acad Med Singapore* 42:687-8.
- World Health Organization and Special Programme for Research and Training in Tropical Diseases Dengue: Guideline for diagnosis, treatment, prevention and control. France: World Health Organization, 2009. Available from: <http://www.who.int/tdr/publications/documents/dengue-diagnosis.pdf>. Accessed: January 2017.